```
   1 CCTGCCTGGT CCTCTGTGCC TGGTGGGGTG GGGGTGCCAG GTGTGTCCAG
  51 AGGAGCCCAT TTGGTAGTGA GGCAGGTATG GGGCTAGAAG CACTGGTGCC
 101 CCTGGCCGTG ATAGTGGCCA TCTTCCTGCT CCTGGTGGAC CTGATGCACC
 151 GGCGCCAACG CTGGGCTGCA CGCTACTCAC CAGGCCCCCT GCCACTGCCC
 201 GGGCTGGGCA ACCTGCTGCA TGTGGACTTC CAGAACACAC CATACTGCTT
 251 CGACCAGTTG CGGCGCCGCT TCGGGGACGT GTTCAGCCTG CAGCTGGCCT
 301 GGACGCCGGT GGTCGTGCTC AATGGGCTGG CGGCCGTGCG CGAGGCGCTG
 351 GTGACCCACG GCGAGGACAC CGCCGACCGC CCGCCTGTGC CCATCACCCA
 401 GATCCTGGGT TTTGGGCCGC GTTCCCAAGG ACGCCCCTTT CGCCCCAACG
 451 GTCTCTTGGA CAAAGCCGTG AGCAACGTGA TCGCCTCCCT CACCTGCGGG
 501 CGCCGCTTCG AGTACGACGA CCCTCGCTTC CTCAGGCTGC TGGACCTAGC
 551 TCAGGAGGGA CTGAAGGAGG AGTCGGGCTT TCTGCGCGAG GTGCTGAATG
 601 CTGTCCCCGT CCTCCTGCAT ATCCCAGCGC TGGCTGGCAA GGTCCTACGC
 651 TTCCAAAAGG CTTTCCTGAC CCAGCTGGAT GAGCTGCTAA CTGAGCACAG
 701 GATGACCTGG GACCCAGCCC AGCCCCCCCG AGACCTGACT GAGGCCTTCC
 751 TGGCAGAGAT GGAGAAGGCC AAGGGGAACC CTGAGAGCAG CTTCAATGAT
 801 GAGAACCTGC GCATAGTGGT GGCTGACCTG TTCTCTGCCG GGATGGTGAC
 851 CACCTCGACC ACGCTGGCCT GGGGCCTCCT GCTCATGATC CTACATCCGG
 901 ATGTGCAGCG CCGTGTCCAA CAGGAGATCG ACGACGTGAT AGGGCAGGTG
 951 CGGCGACCAG AGATGGGTGA CCAGGCTCAC ATGCCCTACA CCACTGCCGT
1001 GATTCATGAG GTGCAGCGCT TTGGGGACAT CGTCCCCCTG GGTGTGACCC
1051 ATATGACATC CCGTGACATC GAAGTACAGG GCTTCCGCAT CCCTAAGGGA
1101 ACGACACTCA TCACCAACCT GTCATCGGTG CTGAAGGATG AGGCCGTCTG
1151 GGAGAAGCCC TTCCGCTTCC ACCCCGAACA CTTCCTGGAT GCCCAGGGCC
1201 ACTTTGTGAA GCCGGAGGCC TTCCTGCCTT TCTCAGCAGG CCGCCGTGCA
1251 TGCCTCGGGG AGCCCCTGGC CCGCATGGAG CTCTTCCTCT TCTTCACCTC
1301 CCTGCTGCAG CACTTCAGCT TCTCGGTGCC CACTGGACAG CCCCGGCCCA
1351 GCCACCATGG TGTCTTTGCT TTCCTGGTGA CCCCATCCCC CTATGAGCTT
1401 TGTGCTGTGC CCCGCTAGAA TGGGGTACCT AGTCCCCAGC CTGCTCCCTA
1451 GCCAGAGGCT CTAATGTACA ATAAAGCAAT GTGGTAGTTC CAAAAAAAAA
1501 AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAA
     (SEQ ID NO: 1)
```

FEATURES:
5'UTR: 1 - 77
Start Codon: 78
Stop Codon: 1416
3'UTR: 1419

Homologous proteins:
Top 10 BLAST Hits

	Score	E
CRA|18000004889269 /altid=gi|181304 /def=gb|AAA53500.1| (M33388...	884	0.0
CRA|18000004927597 /altid=gi|4503223 /def=ref|NP_000097.1| cyto...	883	0.0
CRA|18000004923926 /altid=gi|181306 /def=gb|AAA35737.1| (M33189...	864	0.0
CRA|18000005007118 /altid=gi|2493367 /def=sp|Q29488|CPDH_MACFA ...	827	0.0
CRA|18000005100319 /altid=gi|3913340 /def=sp|O18992|CPDJ_CALJA ...	800	0.0
CRA|18000004884804 /altid=gi|486997 /def=pir||S37284 cytochrome...	682	0.0
CRA|18000004889271 /altid=gi|522195 /def=gb|AAA36403.1| (M24499...	673	0.0
CRA|18000004884803 /altid=gi|461826 /def=sp|Q01361|CPDE_BOVIN C...	669	0.0
CRA|18000004939934 /altid=gi|117244 /def=sp|P13108|CPD4_RAT CYT...	665	0.0
CRA|18000005107537 /altid=gi|2575863 /def=dbj|BAA23125.1| (AB00...	665	0.0

<u>EST:</u>

	Score (bits)	E Value
Sequences producing significant alignments:		
gi|9872134 /dataset=dbest /taxon=960...	775	0.0
gi|6144331 /dataset=dbest /taxon=9606 ...	648	0.0
gi|6703894 /dataset=dbest /taxon=9606 ...	648	0.0

EXPRESSION INFORMATION FOR MODULATORY USE:
gi\|9872134 /liver
gi\|6144331 /kidney
gi\|6703894 /lung

<u>Tissue Expression:</u>

Whole Liver

FIGURE 1

```
  1 MGLEALVPLA VIVAIFLLLV DLMHRRQRWA ARYSPGPLPL PGLGNLLHVD
 51 FQNTPYCFDQ LRRRFGDVFS LQLAWTPVVV LNGLAAVREA LVTHGEDTAD
101 RPPVPITQIL GFGPRSQGRP FRPNGLLDKA VSNVIASLTC GRRFEYDDPR
151 FLRLLDLAQE GLKEESGFLR EVLNAVPVLL HIPALAGKVL RFQKAFLTQL
201 DELLTEHRMT WDPAQPPRDL TEAFLAEMEK AKGNPESSFN DENLRIVVAD
251 LFSAGMVTTS TTLAWGLLLM ILHPDVQRRV QQEIDDVIGQ VRRPEMGDQA
301 HMPYTTAVIH EVQRFGDIVP LGVTHMTSRD IEVQGFRIPK GTTLITNLSS
351 VLKDEAVWEK PFRFHPEHFL DAQGHFVKPE AFLPFSAGRR ACLGEPLARM
401 ELFLFFTSLL QHFSFSVPTG QPRPSHHGVF AFLVTPSPYE LCAVPR
    (SEQ ID NO: 2)
```

FEATURES:
Functional domains and key regions:

[1] PDOC00001 PS00001 ASN_GLYCOSYLATION
N-glycosylation site

```
        347-350 NLSS
```
----
[2] PDOC00005 PS00005 PKC_PHOSPHO_SITE
Protein kinase C phosphorylation site

```
        327-329 TSR
```
----
[3] PDOC00006 PS00006 CK2_PHOSPHO_SITE
Casein kinase II phosphorylation site

```
Number of matches: 5
      1      93-96 THGE
      2    198-201 TQLD
      3    238-241 SFND
      4    327-330 TSRD
      5    437-440 SPYE
```
----
[4] PDOC00008 PS00008 MYRISTYL
N-myristoylation site

```
Number of matches: 2
      1    233-238 GNPESS
      2    255-260 GMVTTS
```
----
[5] PDOC00009 PS00009 AMIDATION
Amidation site

```
Number of matches: 2
      1    140-143 CGRR
      2    387-390 AGRR
```
----
[6] PDOC00081 PS00086 CYTOCHROME_P450
Cytochrome P450 cysteine heme-iron ligand signature

```
        385-394 FSAGRRACLG
```

Membrane spanning structure and domains:

Helix	Begin	End	Score	Certainty
1	3	23	1.877	Certain
2	68	88	1.096	Certain
3	171	191	0.668	Putative
4	252	272	1.914	Certain
5	400	420	1.4C2	Certain
6	425	445	0.833	Putative

FIGURE 2, page 1 of 2

BLAST Alignment to Top Hit:

```
>CRA|18000004889269 /altid=gi|181304 /def=gb|AAA53500.1| (M33388)
          cytochrome P450 IID6 [Homo sapiens] /org=Homo sapiens
          /taxon=9606 /dataset=nraa /length=497
          Length = 497

 Score =  884 bits (2259), Expect = 0.0
 Identities = 444/497 (89%), Positives = 445/497 (89%), Gaps = 51/497 (10%)

Query: 1    MGLEALVPLAVIVAIFLLLVDLMHRRQRWAARYSPGPLPLPGLGNLLHVDFQNTPYCFDQ 60
            MGLEALVPLAVIVAIFLLLVDLMHRRQRWAARY PGPLPLPGLGNLLHVDFQNTPYCFDQ
Sbjct: 1    MGLEALVPLAVIVAIFLLLVDLMHRRQRWAARYPPGPLPLPGLGNLLHVDFQNTPYCFDQ 60

Query: 61   LRRRFGDVFSLQLAWTPVVVLNGLAAVREALVTHGEDTADRPPVPITQILGFGPRSQG-- 118
            LRRRFGDVFSLQLAWTPVVVLNGLAAVREALVTHGEDTADRPPVPITQILGFGPRSQG
Sbjct: 61   LRRRFGDVFSLQLAWTPVVVLNGLAAVREALVTHGEDTADRPPVPITQILGFGPRSQGVF 120

Query: 119  --------------------------------------------------RPFRPNGLLDK 129
                                                              RPFRPNGLLDK
Sbjct: 121  LARYGPAWREQRRFSVSTLRNLGLGKKSLEQWVTEEAACLCAAFANHSGRPFRPNGLLDK 180

Query: 130  AVSNVIASLTCGRRFEYDDPRFLRLLDLAQEGLKEESGFLREVLNAVPVLLHIPALAGKV 189
            AVSNVIASLTCGRRFEYDDPRFLRLLDLAQEGLKEESGFLREVLNAVPVLLHIPALAGKV
Sbjct: 181  AVSNVIASLTCGRRFEYDDPRFLRLLDLAQEGLKEESGFLREVLNAVPVLLHIPALAGKV 240

Query: 190  LRFQKAFLTQLDELLTEHRMTWDPAQPPRDLTEAFLAEMEKAKGNPESSFNDENLRIVVA 249
            LRFQKAFLTQLDELLTEHRMTWDPAQPPRDLTEAFLAEMEKAKGNPESSFNDENLRIVVA
Sbjct: 241  LRFQKAFLTQLDELLTEHRMTWDPAQPPRDLTEAFLAEMEKAKGNPESSFNDENLRIVVA 300

Query: 250  DLFSAGMVTTSTTLAWGLLLMILHPDVQRRVQQEIDDVIGQVRRPEMGDQAHMPYTTAVI 309
            DLFSAGMVTTSTTLAWGLLLMILHPDVQRRVQQEIDDVIGQVRRPEMGDQAHMPYTTAVI
Sbjct: 301  DLFSAGMVTTSTTLAWGLLLMILHPDVQRRVQQEIDDVIGQVRRPEMGDQAHMPYTTAVI 360

Query: 310  HEVQRFGDIVPLGVTHMTSRDIEVQGFRIPKGTTLITNLSSVLKDEAVWEKPFRFHPEHF 369
            HEVQRFGDIVPLGVTHMTSRDIEVQGFRIPKGTTLITNLSSVLKDEAVWEKPFRFHPEHF
Sbjct: 361  HEVQRFGDIVPLGVTHMTSRDIEVQGFRIPKGTTLITNLSSVLKDEAVWEKPFRFHPEHF 420

Query: 370  LDAQGHFVKPEAFLPFSAGRRACLGEPLARMELFLFFTSLLQHFSFSVPTGQPRPSHHGV 429
            LDAQGHFVKPEAFLPFSAGRRACLGEPLARMELFLFFTSLLQHFSFSVPTGQPRPSHHGV
Sbjct: 421  LDAQGHFVKPEAFLPFSAGRRACLGEPLARMELFLFFTSLLQHFSFSVPTGQPRPSHHGV 480

Query: 430  FAFLVTPSPYELCAVPR 446
            FAFLV+PSPYELCAVPR
Sbjct: 481  FAFLVSPSPYELCAVPR 497 (SEQ ID NO: 4)
```

Hmmer search results (Pfam):
Scores for sequence family classification (score includes all domains):

Model	Description	Score	E-value	N
PF00067	Cytochrome P450	516.7	1.7e-151	2

Parsed for domains:

Model	Domain	seq-f	seq-t	hmm-f	hmm-t		score	E-value
PF00067	1/2	35	113	.. 1	92	[.	78.1	2.7e-21
PF00067	2/2	117	443	.. 150	497	.]	442.7	3.3e-129

FIGURE 2, page 2 of 2

```
   1 AGCCTTACAA AGTGCTGGGA TTACCTGCGT GAGCCACCGG GTCCGGCCTC
  51 TTTATGTCTT ACTGTACTGT CTGTCTTGAA AAGTACTTAT TATTTTTGAT
 101 TGGTTCATCA TTTAGTCTAA TTAAAATAAG AGTAGTTTAC ACACCACAAT
 151 TACAGTATTA TAATACTCTG TTTTTCTGTG TGCTTACTAT TACCAGTGAG
 201 TTTTGTACCT TTAGATGATT TCTTCTTGCT CATTAATATC CTTTTTTTTT
 251 TCAGATTGAA AAACTCCCTT TAGCATTTCT TGTGGGATAT AGGTCTGGTG
 301 TTGATGAAAT CTCGCAGCTT TTGTTTGTCT GGGAAGGTCT TTATTTCTCC
 351 TTCCTGTTGG AAGGATATTT TTGCCAGATA CGTTATTCTA GGCTAAAGT
 401 TTTTTTTCCT TCAGCACTTT AAATATGTCA TGCCACTCCC CCCTGGCCTG
 451 TAAGGTTTCC ACTGGAAAGG TGGCTGCCCC ATGTCATGTA TTGGAGCTCT
 501 ACTGCATGTT ATTTGTTTCT TTTCTCTTGC TGCTTTTAGG ATCCTTTCTT
 551 TATCCTTGAC CTTTCGGAGT TTAATTATCA GATGCCTTGA GGTCGTCTTC
 601 TTTGGGTTAA ATCTGCTTGG TGTTCTATAA ACTTCTTGTA CAAAAAATCA
 651 GCCAGGCATG GTGGTGGGCA CCTGTAATCC CAGCTACTTG GGAGGCTGAG
 701 GCAGGAGAAT CGCTTGAACC CTGGAGGTGG AGGTTGCAGT GAGCCGAGAT
 751 CGCATCATTG CACTCCCACC TGGGCGACAG AGCAAAACTC CGTCTCAAAA
 801 AAAAAATTAT TTGGGCTCGG TGGTGCCTGT AGTCCCAGCT ACTTGGGAGG
 851 CAGGAGGTCC ACTTGATGTT GAGATTGCAG TGAGCCATGA TCCTGCCACT
 901 GCACTCCGGC CCGGGCAACA GAGTGAGACC CTGTCTAAAG AAAAAATAAA
 951 AATAAAAAAG CAACATATCC TAAATAAAGG ATCCTCCATA ATGTTTCCAC
1001 CAGATTTCTA ATCAGAAACA TGGAGGCCAG GAAGCAGTGG AGAATGACGA
1051 CCCTCAGGCA GCCCTGGAGG ATGCTGTCAC AGGCTGGGGC AAGGGCCTTC
1101 AGGCTACCAA CTGGGAGCTC TGGGAACAGC CCTGTTGCAA ACAGGAAGTC
1151 ATGGCCCGGC CAGAGCCCAG AATGTGGGCT GAGCTGGGAT CCATGTGACA
1201 GCTTTGAGGC TCACCGGGAG CAGCCTCTGG ACAGGAGAGG TCCCATCCAG
1251 GAAACCTCGG GCATGGCTGG GAAGTGGGGT ACTTGGTGCC GGGTCTGTAT
1301 GTGTGTGTGA CTGGTGTGTG TGAGAGAGAA TGTGTGCCCT GAGTGTCAGT
1351 GTGAGTCTGT GTATGTGTGA ATATTGTCTT TGTGTGGGTG ATTTTCTGCA
1401 TGTGTAATCG TGTCCCTGCA AGTGTGAACA AGTGGACAAG TGTCTGGGAG
1451 TGGACAAGAG ATCTGTGCAC CATCAGGTGT GTGCATAGCG TCTGTGCATG
1501 TCAAGAGTGC AAGGTGAAGT GAAGGGACCA GGCCCATGAT GCCACTCATC
1551 ATCAGGAGCT CTAAGGCCCC AGGTAAGTGC CAGTGACAGA TAAGGGTGCT
1601 GAAGGTCACT CTGGAGTGGG CAGGTGGGGG TAGGGAAAGG GCAAGGTCAT
1651 GTTCTGGAGG AGGGGGTTGTG ACTACATTAG GGTGTATGAG CCTAGCTGGG
1701 AGGTGGATGG CCGGGTCCAC TGAGACCCTG GTTATCCCAG AAGCCTGTGT
1751 GGGCTTGGGG AGCTTGGAGT GGGGAGAGGG GGTGACTTCT CCGACCAGGC
1801 CTTTCTACCA CCCTACCCTG GGTAAGGGCC TGGAGCAGGA AGCAGCGGCA
1851 AGGACCTCTG GAGCAGCCCA TACCTGCCCT GGCCTGACTC TGCCACTGGC
1901 AGCACAGTCA ACACAGCAGG TTCACTCACA GCAGAGGGCG AAGGCCATCA
1951 TCAGCTCCCT TTATAAGGGA AGGGTCACGC GCTCGGTGTG CCGAGAGTGT
2001 CCTGCCTGGT CCTCTGTGCC TGGTGGGGTG GGGGTGCCAG GTGTGTCCAG
2051 AGGAGCCCAG TTGGTAGTGA GGCAGCCATG GGGCTAGAAG CACTGGTGCC
2101 CCTGGCCATG ATAGTGGCCA TCTTCCTGCT CCTGGTGGAC CTGATGCACC
2151 GGCACCAACG CTGGGCTGCA CGCTACCCGC CAGGTCCCCT GCCACTGCCC
2201 GGGCTGGGCA ACCTTGCTGC ATGTGGACTT CCAGAACACA CCATACTGCT
2251 TCGACCAGGT GAGGGAGGAG GTCCTGGAGG GCGGCAGAGG TCCTGAGGAT
2301 GCCCCACCAC CAGCAAACAT GGGTGGTGGG TTAAACCACA GGCTGGATCA
2351 GAAGCCAGGC TGAGAAGGGG AAGCAGGTTT GGGGGACGTT CCTGGGGAAG
2401 GACATTTATA CATGGCATGA AGGACTGGAT TTTCCAAAGG CCAAGGAAGA
2451 GTAGGGCAAG GGCCTGGAGG TGGAGCTGGA CTTGGCAGTG GGCATGCAAG
2501 CCCATTGGGC AACATATGTT ATGGAGTACA AAGTCCCTTC TGCTGACACC
2551 AGAAGGAAAG GCCTTGGCAA TGGAAGATGA GTTAGTCCTG AGTGCCGTTT
2601 AAATCACGAA ATCGAGGATG AAGGGGGTGC AGTGACCCGG TTCAAACCTT
2651 TTGCACTGTG GGTCCTCGGG CCTCACTGCT CACCGGCATG GACCATCATC
2701 TGGGAATGGG ATGCTAACTG GGGCCTCTCG GCAATTTTGG TGACTCTTGC
2751 AAGGTCATAC CTGGGTGACG CATCCAAACT GAGTTCCTCC ATCACAGAAG
2801 GTGTGACCCC CACCCCTGCC CCACGATCAG GAGGCTGGGT CTCCTCCTTC
2851 CACCTGCTCA CTCCTGGTAG CCCCGGGGGT CGTCCAAGGT TCAAATAGGA
2901 CTAGGACCTG TAGTCTGGGG TGATCCTGGC TTGACAAGAG GCCCTGACCC
2951 TCCCTCTGCA GTTGCGGCGC CGCTTCGGGG ACGTGTTCAG CCTGCAGCTG
3001 GCCTGGACGC CGGTGGTCGT GCTCAATGGG CTGGCGGCCG TGCGCGAGGC
3051 GATGGTGACC CGCGGCGAGG ACACGGCCGA CCGCCCGCCT GCGCCCATCT
3101 ACCAGGTCCT GGGCTTCGGG CCGCGTTCCC AAGGCAAGCG GCGGTGGGGG
3151 ACAGAGACCG CGTTTCCGTG GGCCCCGGGT GGACAGTGAC CGTAGCCCAA
3201 GCAGCGCCGA CAGGGCGTGG GGTCCTGGAC GTGAAACAGA GATAAAGGCC
3251 AGCGAGTGGG CTGAGGACAG TGGGCCAGGA AACCACCTGC ACGGGGGAGG
3301 TGCGAGTCTG TGGGCTGGGA GGGGGCGGGG CTACTGCCCA GACCCGCCAG
3351 AAGCCCGGTG GCGAGGCTG ATGCGTCGAA GTGGCGGTGG CGGGGACCGC
3401 GCCTATGCTG CGGGCTCAGT GTGGGCGGGA CGGGCGGGAT CTTCCTTGAG
3451 TGGAAAGGTG GTCAGGGTGG GCAGAGACGA GGTGGGGCCA AACCCCGCCC
3501 CAGGCAGGGG AGCAATGTGG GTGAGCAAAG AGTGGGCCCT GTGCCCAGCT
```

```
3551 GGACCGGGCT AGGGACTGCG GGAGACCTTG TGGAGCGCCA GGGTTGGAGT
3601 GGGTGGCGGA GGGTGGGGCC AAGGCCTTCA TGGCAACGCC CACGTGTCCG
3651 TCCCGCCCCC AGGGGTGATC CTGTCGCGCT ATGGGCCCGC GTGGCGCGAG
3701 CAGAGGCGCT TCTCCGTGTC CACCTTGCGC AACTTGGGCC TGGGCAAGAA
3751 GTCGCTGGAG CAGTGGGTGA CCGAGGAGGC CGCCTGCCTT TGTGCCGCTT
3801 CGCCGACCAA GCCGGTGGCT GATGGGCAGA AGGGCACAAA GCGGGAACTG
3851 GGAAGGCGGG GGACGGAGAA GGCAACCCCT TACCCGCATC TCCCCACCCC
3901 CAGGACGCCC CTTTCGCCCC AACGGCCTCT TGGACAAAGC CGTGAGCAAC
3951 GTGATCGCCT CCCTCACCTG CGGGCGCCGC TTCGAGTACG ACGACCCTCG
4001 CTTCCTCAGG CTGCTGGACC TAGCTCAGGA GGGACTGAAG GAGGAGTCGG
4051 GCTTTCTGCG CGAGGTGCGG AGCGAGAGAC CGAGGAGTCT CTGCAGGGCG
4101 AGCTCCTGAG AGGTGCCGGG GCTGGACTGG GGCCTCCGAA GGGCAGGATT
4151 TGCATAGATG GGTTTGGGAA AGGACATTCC AGGAGACCCC ACTGTAAGAA
4201 GGGCCTGGAG GAGGAGGGGA CATCTCAGAC ATGGTCGTGG GAGAGGTGTG
4251 CCCGGGTCAG GGGGCACCAG GAGAGGCCAA GGACTCTGTA CCCCCGTCCA
4301 CGTTGGAGAT TTCGATTTTA GGTTTCTCCT CTGGGCAAGG AGAGAGGGTG
4351 GAGGCTGGCA CTTGGGGAGG GACTTGGTGA GGTCAGTGGT AAGGACAGGC
4401 AGGCCCTGGG TCTACCTGGA GATGGCTGGG GCCTGAGACT TGTCCAGGTG
4451 AACGCAGAGC ACAGGAGGGA TTGAGACCCC GTTCTGTCTG GTGTAGGTGC
4501 TGAATGCTGT CCCCGTCCTC CTGCACATCC CAGCGCTGGC TGGCAAGGTC
4551 CTACGCTTCC AAAAGGCTTT CCTGACCCAG CTGGATGAGC TGCTAACTGA
4601 GCACAGGATG ACCTGGGACC CAGCCCAGCC ACCCCGAGAC CTGACTGAGG
4651 CCTTCCTGGC AAAGAAGGAG AAGGTGAGAG TGGCTGCCAC GGTGGGGGGC
4701 AAGGGTGGTG GGTTGAACGT CCCAGGAGGA ATGAGGGGAG GCTGGGCAAA
4751 AGGTTGGACC AGTGCATCAC CCGGCGAGCC GCATCTGGGC TGACAGGTGC
4801 AGAATTGGAG GTCATTTGGG GGCTACCCCG TTCTATCCCC TGAGTATCCT
4851 CTCGGCCCTG CTCAGGCCAA GGGGAGCCCT GAGAGCAGCT TCAATGATGA
4901 GAACCTGCGC ATAGTGGTGG GTAACCTGTT CCTTGCCGGG ATGGTGACCA
4951 CCTCGACCAC GCTGGCCTGG GGCCTCCTGC TCATGATCCT ACACCTGGAT
5001 GTGCAGCGTG AGCCCAGCTG GGGCCCAAGG CAGGGACTGA GGGAGGAAGG
5051 GTACAGCTGG GGGCCCCTGG GCTTAGCTGG GACACCCGGG GCTTCCAGCA
5101 CAGGCGTGGC CAGGCTCCTG TAAGCCTAAC TTCCTCCAAC ACAGGAGGAA
5151 GGAGAGTGTC CCCTGGGTGC TGACCCATTG TGGGGACGCA TGTCTGTCCA
5201 GTCCGTGTCC AACAGGAGAT CGACGACGTG ATAGGGCAGG TGCGGCGACC
5251 AGAGATGGGT GACCAGGCTC ACATGCCCTA CACCACTGCC GTGATTCACG
5301 AGGTGCAGCG CTTTGGGGAC ATCATCCCCC TGAGTGTGAC CCATATGACA
5351 TCCCGTGACA TCGAAGTACA GGGCTTCCGC ATCCCTAAGG TAGGCCTGGC
5401 GCCCTCCTCA CCCCAGCTCA GCACCAGCAC CTGGTGATAG CCCCAGCATG
5451 GCTACTGCCA GGTGGGCCCA CTCTAGGAAC CCTGGCCACC TAGTCCTCAA
5501 TGCCACCACA CTGACTGTCC CCACTTGGGT GGGGGGTCCA GAGTATAGGC
5551 AGGGCTGGCC TGTCCATCCA GAGCCCCCGT CTAGTGGGGA GACAAACCAG
5601 GACCTGCCAG AATGTTGGAG GACCCAGCGC CTGCAGGGAG AGGGGGCAGT
5651 GTGGGTGCCT CTGAGAGGTG TGACTGCGCC CTGCTGTGGG GTCGGAGAGG
5701 GTACTGTGGA GCTTCTCGGG CGCAGGACTA GTTGACAGAG TCCAGCTGTG
5751 TGCCAGGCAG TGTGTGTCCC CCGTGTGTTT GGTGGCAGGG GTCCCAGCAT
5801 CCTAGAGTCC AGTCCCCACT CTCACCCTGC ATCTCCTGCC CAGGGAACGA
5851 CACTCATCAC CAACCTGTCA TCGGTGCTGA AGGATGAGGC CGTCTGGGAG
5901 AAGCCCTTCC GCTTCCACCC CGAACACTTC CTGGATGCCC AGGGCCACTT
5951 TGTGAAGCCG GAGGCCTTCC TGCCTTTCTC AGCAGGTGCC TGTGGGGAGC
6001 CCGGCTCCCT GTCCCCTTCC GTGGAGTCTT GCAGGGGTAT CACCCAGGAG
6051 CCAGGCTCAC TGACGCCCCT CCCCTCCCCA CAGGCCGCCG TGCATGCCTC
6101 GGGGAGCCCC TGGCCCGCAT GGAGCTCTTC CTCTTCTTCA CCTCCCTGCT
6151 GCAGCACTTC AGCTTCTCCG TGGCCGCCGG ACAGCCCCGG CCCAGCCACT
6201 CTCGTGTCGT CAGCTTTCTG GTGACCCCAT CCCCCTACGA GCTTTGTGCT
6251 GTGCCCCGCT AGAATGGGGT ACCTAGTCCC CAGCCTGCTC CCTAGCCAGA
6301 GGCTCTAATG TACAATAAAG CAATGTGGTA GTTCCAACTT GGGTCCCCTG
6351 CTCACGCCCT CGTTGGGATC ATCCTCCTCA GGGCAACCCC ACCCCTGCCT
6401 CATTCCTGCT TACCCCACCG CCTGGCCGCA TTTGAGACGG GTACGTTGAG
6451 GCTGAGCAGA TGTCAGTTAC CCTTGCCCAT AATCCCATGT CCCCCACTGA
6501 CCCAACTCTG ACTGCCCAGA TTGGTGACAA GGACTACATT GTCCTGGCAT
6551 GTGGGGAAGG GGCCAGAATG GGCTGACTAG AGGTGTCAGT CAGCCCTGGA
6601 TGTGGTGGAG AGGGCAGCAC TCAGCCTGGA GGCCCATATT TCAGGCCTAA
6651 CTCAGCCCAC CCCACATCAG GGACAGCAGT CCTGCCAGCA CCATCACAAC
6701 AGTCACCTCC CTTCATATAT GACACCCCAA AATGGAAGAC AAATCATGTC
6751 AGGGAGCTAT ATGCCAGGGC TACCTCCCAG GGCTCAGTCG GCAGGTGCCA
6801 GAACATTCCC TGGGAAGGCC CCAGGAAAAC CCAGGACCGA GCCACCGCCC
6851 TCAGCCTGTC ACCTTGTGTC CAAAATTGGT GGGTTCTTGG TCTCACTGAC
6901 TTCAAGAATG AAGCCGTGGA CCCTCACGGT GAGTGTTACA GTTCTTAAAG
6951 ATGGTGTGTT CAGAGTTTGT TCCTTCTGAT GTTAAGACGT GTTCAGAGTT
7001 TCTTCCTTCT GGTGGGTGCG TGGTCTTGCT GGCTTCAGGA GTGAAGCTGC
7051 AGACCTTCAC AGTGAGTGTT ACGGCTCTTA AGGCTGCACG TACGGAGTTG
```

FIGURE 3, page 2 of 5

```
7101 TTCATTCTTC CTGGTGGGTT TGTGGTCTCA CTGGCCTCAG GAGTGAAACT
7151 GCAGTCCTTC CAGTGTTACA ACTCATAAAG GCAGTGTGGA CCCAATGAGG
7201 GAGCAGCAGC AGCAAGACTT ACTGCAAACA GCAAAAGAAT GATGGCAACC
7251 AGGTTGCCGC TGCTACTTCA GGCAGCCTGC TTTTATTCCC TTATCTGACC
7301 CCCACCCACA TCCTGCTGAT TGGCCCATTT TACAGACAGT GGATTGGTCC
7351 ACTTACAGAG AGCTGATTGG TGCATTTACA ATCCCTGAGC TAGACACAGA
7401 GTACTGATTG GTATATTTAC AAACCTTGAG CTAGACACAG AGTGCTGAAT
7451 GGTGTATTTA CAATCCCTTA GCTAGACATA AAGGTTGTCC CAGTCCCCAC
7501 TAGATTAGCT AGATAGAGTA GACAGAGAGC ACTGATTGGT GCGTTTACAA
7551 ACCTTGAGTT AGACACAGGG TGCTGACTGG TGTGTTTACA AACCTTGAGC
7601 TAGACACAGA GTGCTGATTG GTGTATTTAC AATCTTTTAG CTAGAAATAA
7651 AGGTTCCCCA AGTCCCCACC AGATTAGCTA GATAGAGTGC TAATTGGTGC
7701 ATGCACGAAC CCGGAGCTAG ACACAGAGTG CTGATTGGTG CATATACAAT
7751 CCTCTGGCTA GACATAAAAG TTCTCCAAGT CCCCACCTGA CTCAGGAGCC
7801 CAGCCAGCTT CGCCTAGTGG ATCCTATGCC AGGGCCACAG GCAGAGCTGC
7851 CTGCTAGTCC CACACCGGGC ACCTGTACTC CTCAGCCCTT GGGCAGTGGA
7901 CGGGACCAGG TGCCGTGGAG CAGTGGGAGG CACCCATCCG GGAGGCTCGG
7951 GCCTCGCAGG GAGCCCACCG TAGGGAGGCT TGGGCATGGC AGGCTGCAAG
8001 TCCTGAGCCC TGCCCCGCGG GGAGGTGACT GAGGCCTGGC GACAATTCAA
8051 GTGTGGTGAG CGCCGGCAGG CCAGCAGTAC TGGGGGACCC GGTGCCCCCT
8101 CTGCAGCTGC TGGCCCAGGT GCTAAGCCCC TCACTGCCTG GGGCCAGAGG
8151 CACCAGCCGG CCGCTCCGAG TGCAGGGCCC GCTGAGCCCC TGCCCACCCA
8201 GAACTGGTGC TGGCCCGCGA GCAACCCAGG TTCCCGCACA CGCCTCTCCC
8251 TCCATACCTC CCCGCAAGCA GACGGAGCCG GCTCCAGCCT CCACCAGTCC
8301 AGAGAGGGGC TCCCACAGTG CAGCGCTGGG CTGAACAAGG TCCTACGCTT
8351 CCAAAAGGCT TTCCTGACCC AGCTGGATGA GCTGCTAACT GAGCACAGGA
8401 TGACCTGGGA CCCAGCCCAG CCCCCCCGAG ACCTGACTGA GGCCTTTCCT
8451 GGCAGAGATG GAGAAGGTGA GAGTGGCTGC CACGGTGGGG GGCAAGGGTG
8501 GTGGGTTGAG CGTCCCAGGA GGAATGAGGG GAGGCTGGGC AAAAGGTTGG
8551 ACCAGTGCAT CACCCGGCGA GCCGCATCTG GGCTGACAGG TGCAGAATTG
8601 GAGGTCATTT GGGGGCTACC CCGTTCTGTC CCGAGTATGC TCTCGGCCCT
8651 GCTCAGGCCA AGGGGAACCC TGAGAGCAGC TTCAATGATG AGAACCTGCG
8701 CATAGTGGTG GCTGACCTGT TCTCTGCCGG GATGGTGACC ACCTCGACCA
8751 CGCTGGCCTG GGGCCTCCTG CTCATGATCC TACATCCGGA TGTGCAGCGT
8801 GAGCCCATCT GGGAAACAGT GCAGGGGCCG AGGGAGGAAG GGTACAGGCG
8851 GGGGCCCATG AACTTTGCTG GGACACCCGG GGCTCCAAGC ACAGGCTTGA
8901 CCAGGATCCT GTAAGCCTGA CCTCCTCCAA CATAGGAGGC AAGAAGGAGT
8951 GTCAGGGCCG GACCCCCTGG GTGCTGACCC ATTGTGGGGA CGCATGTCTG
9001 TCCAGGCCGT GTCCAACAGG AGATCGACGA CGTGATAGGG CAGGTGCGGC
9051 GACCAGAGAT GGGTGACCAG GCTCACATGC CCTACACCAC TGCCGTGATT
9101 CATGAGGTGC AGCGCTTTGG GGACATCGTC CCCCTGGGTG TGACCCATAT
9151 GACATCCCGT GACATTCGAA GTACAGGGCT TCCGCATCCC TAAGGTAGGC
9201 CTGGCGCCNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
9251 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
9301 NNNNNNCCT GCCCAGGGAA CGACACTCAT CACCAACCTG TCATCGGTGC
9351 TGAAGGATGA GGCCGTCTGG GAGAAGCCCT TCCGCTTCCA CCCCGAACAC
9401 TTCCTGGATG CCCAGGGCCA CTTTGTGAAG CCGGAGGCCT TCCTGCCTTT
9451 CTCAGCAGGT GCCTGTGGGG AGCCCGGCTC CCTGTCCCCT TCCGTGGAGT
9501 CTTGCAGGGG TATCACCCAG GAGCCAGGCT CACTGACGCC CCTCCCCTCC
9551 CCACAGGCCG CCGTGCATGC CTCGGGGAGC CCCTGGCCCG CATGGAGCTC
9601 TTCCTCTTCT TCACCTCCCT GCTGCAGCAC TTCAGCTTCT CGGTGCCCAC
9651 TGGACAGCCC CGGCCCAGCC ACCATGGTGT CTTTGCTTTC CTGGTGAGCC
9701 CATCCCCCTA TGAGCTTTGT GCTGTGCCCC GCTAGAATGG GGTACCTAGT
9751 CCCCAGCCTG CTCCCTAGCC AGAGGCTCTA ATGTACAATA AAGCAATGTG
9801 GTAGTTCCAA CTCGGGTCCC CTGCTCACGC CCTCGTTGGG ATCATCCTCC
9851 TCAGGGCAAC CCCACCCCTG CCTCATTCCT GCTTACCCCA CCGCCTGGCC
9901 GCATTTGAGA CAGGGGTACG TTGAGGCTGA GCAGATGTCA GTTACCCTTG
9951 CCCATAATCC CATGTCCCCC ACTGACCCAA CTCTGACTGC CCAGATTGGT
10001 GACAAGGACT ACATTGTCCT GGCATGTGGG GAAGGGGCCA GAATGGGCTG
10051 ACTAGAGGTG TCAGTCAGCC CTGGATGTGG TGGAGAGGGC AGGACTCAGC
10101 CTGGAGGCCC ATATTTCAGG CCTAACTCAG CCCACCCCAC ATCAGGGACA
10151 GCAGTCCTGC CAGCACCATC ACAACAGTCA CCTCCCTTCA TATATGACAC
10201 CCCAAAACGG AAGACAAATC ATGGCGTCAG GGAGCTATAT GCCAGGGCTA
10251 CCTACCTCCC AGGGCTCAGT CGGCAGGT
     (SEQ ID NO: 3)

FEATURES:
Start........2078
Exon:    2078-2258
Intron:  2259-2961
Exon:    2962-3133
```

FIGURE 3, page 3 of 5

```
Intron:   3134-3903
Exon:     3904-4064
Intron:   4065-4496
Exon:     4497-4673
Intron:   4674-4865
Exon:     4866-5007
Intron:   5008-5201
Exon:     5202-5389
Intron:   5390-5843
Exon:     5844-5985
Intron:   5986-9556
Exon:     9557-9732
Stop      9733
```

SNPs:

DNA Position	Major	Minor	Domain	Protein Position	Major	Minor
3101	C	T A	Exon	107	T	T T
3439	A	G	Intron			
4908	C	T	Exon	245	P	L
5627	G	A	Intron			
6733	T	C	Intron			
7788	–	C T	Intron			
7867	G	A	Intron			
7948	C	T	Intron			

Context:

DNA
Position

```
3101     GTGTGACCCCCACCCCTGCCCCACGATCAGGAGGCTGGGTCTCCTCCTTCCACCTGCTCA
         CTCCTGGTAGCCCCGGGGGTCGTCCAAGGTTCAAATAGGACTAGGACCTGTAGTCTGGGG
         TGATCCTGGCTTGACAAGAGGCCCTGACCCTCCCTCTGCAGTTGCGGCGCCGCTTCGGGG
         ACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTGCTCAATGGGCTGGCGGCCG
         TGCGCGAGGCGATGGTGACCCGCGGCGAGGACACGGCCGACCGCCCGCCTGCGCCCATCT
         [C,T,A]
         CCAGGTCCTGGGCTTCGGGCCGCGTTCCCAAGGCAAGCGGCGGTGGGGGGACAGAGACCGC
         GTTTCCGTGGGCCCCGGGTGGACAGTGACCGTAGCCCAAGCAGCGCCGACAGGGCGTGGG
         GTCCTGGACGTGAAACAGAGATAAAGGCCAGCGAGTGGGCTGAGGACAGTGGGCCAGGAA
         ACCACCTGCACGGGGGGAGGTGCGAGTCTGTGGGCTGGGAGGGGGCGGGGCTACTGCCCAG
         ACCCGCCAGAAGCCCGGTGGGCGAGGCTGATGCGTCGAAGTGGCGGTGGCGGGGGACCGCG

3439     CGGCGGTGGGGGGACAGAGACCGCGTTTCCGTGGGCCCCGGGTGGACAGTGACCGTAGCCC
         AAGCAGCGCCGACAGGGCGTGGGGGTCCTGGACGTGAAACAGAGATAAAGGCCAGCGAGTG
         GGCTGAGGACAGTGGGCCAGGAAACCACCTGCACGGGGGGAGGTGCGAGTCTGTGGGCTGG
         GAGGGGGCGGGGCTACTGCCCAGACCCGCCAGAAGCCCGGTGGGCGAGGCTGATGCGTCG
         AAGTGGCGGTGGCGGGGGACCGCGCCTATGCTGCGGGCTCAGTGTGGGCGGGGACGGGCGGG
         [A,G]
         TCTTCCTTGAGTGGAAAGGTGGTCAGGGTGGGCAGAGACGAGGTGGGGGCCAAACCCCGCC
         CCAGGCAGGGGAGCAATGTGGGTGAGCAAAGAGTGGGCCCTGTGCCCAGCTGGACCGGGC
         TAGGGACTGCGGGAGACCTTGTGGAGCGCCAGGGTTGGAGTGGGTGGCGGAGGGTGGGGC
         CAAGGCCTTCATGGCAACGCCCACGTGTCCGTCCCGCCCCCAGGGGTGATCCTGTCGCGC
         TATGGGCCCGCGTGGCGCGAGCAGAGGCGCTTCTCCGTGTCCACCTTGCGCAACTTGGGC

4908     ATGACCTGGGACCCAGCCCAGCCACCCCGAGACCTGACTGAGGCCTTCCTGGCAAAGAAG
         GAGAAGGTGAGAGTGGCTGCCACGGTGGGGGGGCAAGGGTGGTGGGTTGAACGTCCCAGGA
         GGAATGAGGGGAGGCTGGGCAAAAGGTTGGACCAGTGCATCACCCGGCGAGCCGCATCTG
         GGCTGACAGGTGCAGAATTGGAGGTCATTTGGGGGCTACCCCGTTCTATCCCCTGAGTAT
         CCTCTCGGCCCTGCTCAGGCCAAGGGGAGCCCTGAGAGCAGCTTCAATGATGAGAACCTG
         [C,T]
         GCATAGTGGTGGGTAACCTGTTCCTTGCCGGGATGGTGACCACCTCGACCACGCTGGCCT
         GGGGCCTCCTGCTCATGATCCTACACCTGGATGTGCAGCGTGAGCCCAGCTGGGGCCCAA
         GGCAGGGACTGAGGGAGGAAGGGTACAGCTGGGGGCCCCTGGGCTTAGCTGGGACACCCG
         GGGCTTCCAGCACAGGCGTGGCCAGGCTCCTGTAAGCCTAACTTCCTCCAACACAGGAGG
         AAGGAGAGTGTCCCCTGGGTGCTGACCCATTGTGGGGACGCATGTCTGTCCAGTCCGTGT

5627     CCCCTGAGTGTGACCCATATGACATCCCGTGACATCGAAGTACAGGGCTTCCGCATCCCT
```

FIGURE 3, page 4 of 5

```
        AAGGTAGGCCTGGCGCCCTCCTCACCCCAGCTCAGCACCAGCACCTGGTGATAGCCCCAG
        CATGGCTACTGCCAGGTGGGCCCACTCTAGGAACCCTGGCCACCTAGTCCTCAATGCCAC
        CACACTGACTGTCCCCACTTGGGTGGGGGGTCCAGAGTATAGGCAGGGCTGGCCTGTCCA
        TCCAGAGCCCCCGTCTAGTGGGGAGACAAACCAGGACCTGCCAGAATGTTGGAGGACCCA
        [G,A]
        CGCCTGCAGGGAGAGCGGGCAGTGTGGGTGCCTCTGAGAGGTGTGACTGCGCCCTGCTGT
        GGGGTCGGAGAGGGTACTGTGGAGCTTCTCGGGCGCAGGACTAGTTGACAGAGTCCAGCT
        GTGTGCCAGGCAGTGTGTGTCCCCCGTGTGTTTGGTGGCAGGGGTCCCAGCATCCTAGAG
        TCCAGTCCCCACTCTCACCCTGCATCTCCTGCCCAGGGAACGACACTCATCACCAACCTG
        TCATCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACAC

6733    TGAGACGGGTACGTTGAGGCTGAGCAGATGTCAGTTACCCTTGCCCATAATCCCATGTCC
        CCCACTGACCCAACTCTGACTGCCCAGATTGGTGACAAGGACTACATTGTCCTGGCATGT
        GGGGAAGGGGCCAGAATGGGCTGACTAGAGGTGTCAGTCAGCCCTGGATGTGGTGGAGAG
        GGCAGGACTCAGCCTGGAGGCCCATATTTCAGGCCTAACTCAGCCCACCCCACATCAGGG
        ACAGCAGTCCTGCCAGCACCATCACAACAGTCACCTCCCTTCATATATGACACCCCAAAA
        [T,C]
        GGAAGACAAATCATGTCAGGGAGCTATATGCCAGGGCTACCTCCCAGGGCTCAGTCGGCA
        GGTGCCAGAACATTCCCTGGGAAGGCCCCAGGAAAACCCAGGACCGAGCCACCGCCCTCA
        GCCTGTCACCTTGTGTCCAAAATTGGTGGGTTCTTGGTCTCACTGACTTCAAGAATGAAG
        CCGTGGACCCTCACGGTGAGTGTTACAGTTCTTAAAGATGGTGTGTTCAGAGTTTGTTCC
        TTCTGATGTTAAGACGTGTTCAGAGTTTCTTCCTTCTGGTGGGTGCGTGGTCTTGCTGGC

7788    TCCCAGTCCCCACTAGATTAGCTAGATAGAGTAGACAGAGAGCACTGATTGGTGCGTTTA
        CAAACCTTGAGTTAGACACAGGGTGCTGACTGGTGTGTTTACAAACCTTGAGCTAGACAC
        AGAGTGCTGATTGGTGTATTTACAATCTTTTAGCTAGAAATAAAGGTTCCCCAAGTCCCC
        ACCAGATTAGCTAGATAGAGTGCTAATTGGTGCATGCACGAACCCGGAGCTAGACACAGA
        GTGCTGATTGGTGCATATACAATCCTCTGGCTAGACATAAAAGTTCTCCAAGTCCCCACC
        [-,C,T]
        GACTCAGGAGCCCAGCCAGCTTCGCCTAGTGGATCCTATGCCAGGGCCACAGGCAGAGCT
        GCCTGCTAGTCCCACACCGGGCACCTGTACTCCTCAGCCCTTGGGCAGTGGACGGGACCA
        GGTGCCGTGGAGCAGTGGGAGGCACCCATCCGGGAGGCTCGGGCCTCGCAGGGAGCCCAC
        CGTAGGGAGGCTTGGGCATGGCAGGCTGCAAGTCCTGAGCCCTGCCCCGCGGGGAGGTGA
        CTGAGGCCTGGCGACAATTCAAGTGTGGTGAGCGCCGGCAGGCCAGCAGTACTGGGGGAC

7867    AGGGTGCTGACTGGTGTGTTTACAAACCTTGAGCTAGACACAGAGTGCTGATTGGTGTAT
        TTACAATCTTTTAGCTAGAAATAAAGGTTCCCCAAGTCCCCACCAGATTAGCTAGATAGA
        GTGCTAATTGGTGCATGCACGAACCCGGAGCTAGACACAGAGTGCTGATTGGTGCATATA
        CAATCCTCTGGCTACACATAAAAGTTCTCCAAGTCCCCACCTGACTCAGGAGCCCAGCCA
        GCTTCGCCTAGTGGATCCTATGCCAGGGCCACAGGCAGAGCTGCCTGCTAGTCCCACACC
        [G,A]
        GGCACCTGTACTCCTCAGCCCTTGGGCAGTGGACGGGACCAGGTGCCGTGGAGCAGTGGG
        AGGCACCCATCCGGGAGGCTCGGGCCTCGCAGGGAGCCCACCGTAGGGAGGCTTGGGCAT
        GGCAGGCTGCAAGTCCTGAGCCCTGCCCCGCGGGGAGGTGACTGAGGCCTGGCGACAATT
        CAAGTGTGGTGAGCGCCGGCAGGCCAGCAGTACTGGGGGACCCGGTGCCCCCTCTGCAGC
        TGCTGGCCCAGGTGCTAAGCCCCTCACTGCCTGGGGCCAGAGGCACCAGCCGGCCGCTCC

7948    TAAAGGTTCCCCAAGTCCCCACCAGATTAGCTAGATAGAGTGCTAATTGGTGCATGCACG
        AACCCGGAGCTAGACACAGAGTGCTGATTGGTGCATATACAATCCTCTGGCTAGACATAA
        AAGTTCTCCAAGTCCCCACCTGACTCAGGAGCCCAGCCAGCTTCGCCTAGTGGATCCTAT
        GCCAGGGCCACAGGCAGAGCTGCCTGCTAGTCCCACACCGGGCACCTGTACTCCTCAGCC
        CTTGGGCAGTGGACGGGACCAGGTGCCGTGGAGCAGTGGGAGGCACCCATCCGGGAGGCT
        [C,T]
        GGGCCTCGCAGGGAGCCCACCGTAGGGAGGCTTGGGCATGGCAGGCTGCAAGTCCTGAGC
        CCTGCCCCGCGGGGAGGTGACTGAGGCCTGGCGACAATTCAAGTGTGGTGAGCGCCGGCA
        GGCCAGCAGTACTGGGGGACCCGGTGCCCCCTCTGCAGCTGCTGGCCCAGGTGCTAAGCC
        CCTCACTGCCTGGGGCCAGAGGCACCAGCCGGCCGCTCCGAGTGCAGGGCCCGCTGAGCC
        CCTGCCCACCCAGAACTGGTGCTGGCCCGCGAGCAACCCAGGTTCCCGCACACGCCTCTC
```

Chromosome mapping:
Chromosome #22

FIGURE 3, page 5 of 5